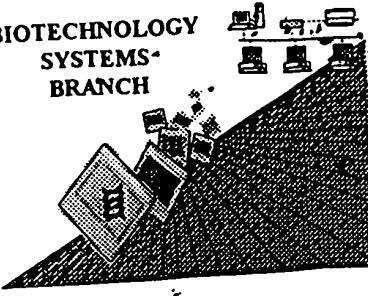


BIOTECHNOLOGY  
SYSTEMS  
BRANCH

RAW SEQUENCE LISTING  
ERROR REPORT



The Biotechnology Systems Branch of the Scientific and Technical Information Center (STIC) detected errors when processing the following computer readable form:

Application Serial Number: 09/936921  
Source: PCT 09  
Date Processed by STIC: 10/04/01

THE ATTACHED PRINTOUT EXPLAINS DETECTED ERRORS.  
PLEASE FORWARD THIS INFORMATION TO THE APPLICANT BY EITHER:  
1) INCLUDING A COPY OF THIS PRINTOUT IN YOUR NEXT COMMUNICATION TO THE  
APPLICANT, WITH A NOTICE TO COMPLY or,  
2) TELEPHONING APPLICANT AND FAXING A COPY OF THIS PRINTOUT, WITH A  
NOTICE TO COMPLY  
FOR CRF SUBMISSION QUESTIONS, PLEASE CONTACT MARK SPENCER, 703-308-4212.

FOR SEQUENCE RULES INTERPRETATION, PLEASE CONTACT ROBERT WAX, 703-308-4216.  
PATENTIN 2.1 e-mail help: [patin21help@uspto.gov](mailto:patin21help@uspto.gov) or phone 703-306-4119 (R. Wax)  
PATENTIN 3.0 e-mail help: [patin3help@uspto.gov](mailto:patin3help@uspto.gov) or phone 703-306-4119 (R. Wax)

TO REDUCE ERRORED SEQUENCE LISTINGS, PLEASE USE THE CHECKER  
VERSION 3.0 PROGRAM, ACCESSIBLE THROUGH THE U.S. PATENT AND  
TRADEMARK OFFICE WEBSITE. SEE BELOW:

**Checker Version 3.0**

The Checker Version 3.0 application is a state-of-the-art Windows based software program employing a logical and intuitive user-interface to check whether a sequence listing is in compliance with format and content rules. Checker Version 3.0 works for sequence listings generated for the original version of 37 CFR §§1.821 – 1.825 effective October 1, 1990 (old rules) and the revised version (new rules) effective July 1, 1998 as well as World Intellectual

Property Organization (WIPO) Standard ST.25.  
Checker Version 3.0 replaces the previous DOS-based version of Checker, and is Y2K-compliant. Checker allows public users to check sequence listings in Computer Readable form (CRF) before submitting them to the United States Patent and Trademark Office (USPTO). Use of Checker prior to filing the sequence listing is expected to result in fewer errored sequence listings, thus saving time and money.

Checker Version 3.0 can be down loaded from the USPTO website at the following address:  
<http://www.uspto.gov/web/offices/pac/checker>

Raw Sequence Listing Error Summary

SERIAL NUMBER: 09/936921

ERROR DETECTED

SUGGESTED CORRECTION

ATTN: NEW RULES CASES: PLEASE DISREGARD ENGLISH "ALPHA" HEADERS, WHICH WERE INSERTED BY PTO SOFT

1  Wrapped Nucleic  
Wrapped Aminos

The number/text at the end of each line "wrapped" down to the next line. This may occur if your file was retrieved in a word processor after creating it. Please adjust your right margin to .3; this will prevent "wrapping."

2  Invalid Line Length

The rules require that a line not exceed 72 characters in length. This includes white spaces.

3  Misaligned Amino  
Numbering

The numbering under each 5<sup>th</sup> amino acid is misaligned. Do not use tab codes between numbers; use space characters, instead.

4  Non-ASCII

The submitted file was not saved in ASCII(DOS) text, as required by the Sequence Rules. Please ensure your subsequent submission is saved in ASCII text.

5  Variable Length

Sequence(s)  contain n's or Xaa's representing more than one residue. Per Sequence Rules, each n or Xaa can only represent a single residue. Please present the maximum number of each residue having variable length and indicate in the <220>-<223> section that some may be missing.

6  PatentIn 2.0  
"bug"

A "bug" in PatentIn version 2.0 has caused the <220>-<223> section to be missing from amino acid sequences(s)  . Normally, PatentIn would automatically generate this section from the previously coded nucleic acid sequence. Please manually copy the relevant <220>-<223> section to the subsequent amino acid sequence. This applies to the mandatory <220>-<223> sections for Artificial or Unknown sequences.

7  Skipped Sequences  
(OLD RULES)

Sequence(s)  missing. If intentional, please insert the following lines for each skipped sequence:  
(2) INFORMATION FOR SEQ ID NO:X: (insert SEQ ID NO where "X" is shown)  
(i) SEQUENCE CHARACTERISTICS: (Do not insert any subheadings under this heading)  
(ii) SEQUENCE DESCRIPTION:SEQ ID NO:X: (insert SEQ ID NO where "X" is shown)  
This sequence is intentionally skipped

8  Skipped Sequences  
(NEW RULES)

Please also adjust the "(ii) NUMBER OF SEQUENCES:" response to include the skipped sequences.

9  Use of n's or Xaa's  
(NEW RULES)

Sequence(s)  missing. If Intentional, please insert the following lines for each skipped sequence.  
<210> sequence id number  
<400> sequence id number  
000

10  Invalid <213>  
Response

Per 1.823 of Sequence Rules, the only valid <213> responses are: Unknown, Artificial Sequence, or scientific name (Genus/species). <220>-<223> section is required when <213> response is Unknown or is Artificial Sequence

11  Use of <220>

Sequence(s)  missing the <220> "Feature" and associated numeric identifiers and responses. Use of <220> to <223> is MANDATORY if <213> "Organism" response is "Artificial Sequence" or "Unknown." Please explain source of genetic material in <220> to <223> section.  
(See "Federal Register," 06/01/1998, Vol. 63, No. 104, pp. 29631-32) (Sec. 1.823 of Sequence Rules)

12  PatentIn 2.0  
"bug"

Please do not use "Copy to Disk" function of PatentIn version 2.0. This causes a corrupted file, resulting in missing mandatory numeric identifiers and responses (as indicated on raw sequence listing). Instead, please use "File Manager" or any other manual means to copy file to floppy disk.

13  Misuse of n

n can only be used to represent a single nucleotide in a nucleic acid sequence. N is not used to represent any value not specifically a nucleotide.

PCT09

RAW SEQUENCE LISTING  
 PATENT APPLICATION: US/09/936,921

DATE: 10/04/2001  
 TIME: 10:50:38

Input Set : A:\seqlisting110530\_pac.txt  
 Output Set: N:\CRF3\10042001\I936921.raw

```

3 <110> APPLICANT: RAOULT, Didier
4           LA SCOLA, Bernard
5           BIRG, Marie-Laure
6           FENOLLAR, Florence
8 <120> TITLE OF INVENTION: Diagnosis of Whipple's Disease
10 <130> FILE REFERENCE: 110530
C--> 12 <140> CURRENT APPLICATION NUMBER: US/09/936,921
C--> 12 <141> CURRENT FILING DATE: 2001-09-20
12 <150> PRIOR APPLICATION NUMBER: PCT/FR00/00754
13 <151> PRIOR FILING DATE: 2000-03-24
15 <150> PRIOR APPLICATION NUMBER: FR99/06679
16 <151> PRIOR FILING DATE: 1999-05-21
18 <150> PRIOR APPLICATION NUMBER: FR99/03989
19 <151> PRIOR FILING DATE: 1999-03-26
21 <160> NUMBER OF SEQ ID NOS: 5
23 <170> SOFTWARE: PatentIn version 3.0
25 <210> SEQ ID NO: 1
26 <211> LENGTH: 19
27 <212> TYPE: DNA
C--> 28 <213> ORGANISM: Artificial
30 <220> FEATURE:
31 <223> OTHER INFORMATION: primer
33 <220> FEATURE:
34 <221> NAME/KEY: modified_base
35 <222> LOCATION: (2)..(2)
36 <223> OTHER INFORMATION: i
39 <220> FEATURE:
40 <221> NAME/KEY: modified_base
41 <222> LOCATION: (8)..(9)
42 <223> OTHER INFORMATION: i
45 <220> FEATURE:
46 <221> NAME/KEY: modified_base
47 <222> LOCATION: (11)..(11)
48 <223> OTHER INFORMATION: i
51 <400> SEQUENCE: 1
W--> 52 tnatgggnnc naanatgca
55 <210> SEQ ID NO: 2
56 <211> LENGTH: 20
57 <212> TYPE: DNA
C--> 58 <213> ORGANISM: Artificial
60 <220> FEATURE:
61 <223> OTHER INFORMATION: primer
63 <220> FEATURE:
64 <221> NAME/KEY: modified_base
65 <222> LOCATION: (6)..(6)
66 <223> OTHER INFORMATION: i
69 <220> FEATURE:

```

Does Not Comply  
 Corrected Diskette Needed

See page 1 of 4

And Error Summary Sheet  
 item nine.

*Erroneous location 14 in the sequence  
 represented as an unknown is not  
 identified in field 21, 22 or 23. ↗*

*The type of errors shown may  
 exist throughout the Sequence Listing. Please check subsequent  
 sequences for similar errors.*

19

**RAW SEQUENCE LISTING**  
**PATENT APPLICATION: US/09/936,921**

DATE: 10/04/2001  
 TIME: 10:50:38

Input Set : A:\seqlisting110530\_pac.txt  
 Output Set: N:\CRF3\10042001\I936921.raw

```

70 <221> NAME/KEY: modified_base
71 <222> LOCATION: (15)..(15)
72 <223> OTHER INFORMATION: i
75 <220> FEATURE:
76 <221> NAME/KEY: modified_base
77 <222> LOCATION: (18)..(18)
78 <223> OTHER INFORMATION: i
81 <400> SEQUENCE: / 20
W--> 82 gcccancatt ccatntchcc
85 <210> SEQ ID NO: 3
86 <211> LENGTH: 612
87 <212> TYPE: DNA
88 <213> ORGANISM: Tropheryma whippelii
89 <220> FEATURE:
90 <221> NAME/KEY: misc_feature
91 <222> LOCATION: (1)..(612) ✓
92 <223> OTHER INFORMATION: n=a or g or c or t/u, unknown, or other.
93 <400> SEQUENCE: 3
W--> 97 aagtccccng gacggnanag gnatggagng gtatgtancg atcgatgcgg gtgatgtttt 60
W--> 99 aattgccnag gatccggca ttgtggggga tgttccgct gatgttgtca ctgtcannca 120
101 ggatgacggg aaacatcgcg actaccatgt tggtaaattt gtcgttcaa atcaggggcaa 180
W--> 103 ctgttacaaac cagnagttt tggtccgatc cggagatcggt gtataaaaag gtacagttct 240
105 tgcacatgtt ccattactg acaaagggtt gcttagtctt ggttagaaatc ttctgggtgc 300
107 tttcatgccc tgggagggtt ataacttta ggtatgcgata attatcagcc agaatttggt 360
W--> 109 caaggacgac accctttcnt caatccacat agaagaacat gaggttagca cccgggatac 420
W--> 111 gaagctgggc agtganagaa ataacgcgag accttccgaa tgtaagcatg gattacataa 480
113 aggacttggg cgaacgggtt attatccgga ttggcgctga ggttggccct gggacattt 540
115 tggttgttaa ggtgacccca aaggcgaag accgaactca acgcggaaga acgtttgtg 600
117 agggctatct tt 612
120 <210> SEQ ID NO: 4
121 <211> LENGTH: 18
122 <212> TYPE: DNA
C--> 123 <213> ORGANISM: Artificial
125 <220> FEATURE:
126 <223> OTHER INFORMATION: oligonucleotide
128 <400> SEQUENCE: 4 18
129 gcattgtggg ggatgttt
132 <210> SEQ ID NO: 5
133 <211> LENGTH: 18
134 <212> TYPE: DNA
C--> 135 <213> ORGANISM: Artificial
137 <220> FEATURE:
138 <223> OTHER INFORMATION: oligonucleotide
140 <400> SEQUENCE: 5 18
141 ttggggtcac cttaccaa

```

VERIFICATION SUMMARY  
PATENT APPLICATION: US/09/936,921

DATE: 10/04/2001  
TIME: 10:50:39

Input Set : A:\seqlisting110530\_pac.txt  
Output Set: N:\CRF3\10042001\I936921.raw

L:12 M:270 C: Current Application Number differs, Replaced Current Application No  
L:12 M:271 C: Current Filing Date differs, Replaced Current Filing Date  
L:28 M:220 C: Keyword misspelled or invalid format, <213> ORGANISM for SEQ ID#:1  
L:52 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:1  
L:58 M:220 C: Keyword misspelled or invalid format, <213> ORGANISM for SEQ ID#:2  
L:82 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:2  
L:97 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:3  
L:99 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:3  
L:103 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:3  
L:109 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:3  
L:111 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:3  
L:123 M:220 C: Keyword misspelled or invalid format, <213> ORGANISM for SEQ ID#:4  
L:135 M:220 C: Keyword misspelled or invalid format, <213> ORGANISM for SEQ ID#:5